CYREN

Fighting Spam, Phishing and Malware With Recurrent Pattern Detection

White Paper — September 2017

Fighting Spam, Phishing and Malware With Recurrent Pattern Detection

Spam, phishing and email-borne malware such as viruses and worms are most often released in large quantities in a relatively short period of time, causing global damage in the tens of billions of dollars annually.

Although many methods to combat these threats have been developed throughout the past several years, a common deficiency of most of these methods is that they lack the ability to adapt quickly enough to the rapid change of distribution and infiltration techniques invented by spammers and malware authors. The objective of this document is to discuss the characteristics of these threats, the challenges facing technologies that aim to mitigate these attacks, and describe how the Cyren solution protects against all types of email-borne threats.

The challenges of accurate detection

Detecting spam, phishing and malware presents a range of challenges:

- Spam Detection: In composing spam messages, spammers use sophisticated tactics to evade existing spam detection applications. This includes covering the tracks to the spammers, manipulating or hiding the commercial URLs, use of non-English words and phrases and a host of other methods. Typically a spam outbreak will only last a few hours and be launched from a network of zombie machines. To complicate the detection process, each message within the spam outbreak can be composed differently and employ more than one evasion technique.
- Phishing Detection: Phishing messages appear to be from genuine or credible sources. Like spam, phishing messages can be sent in any language or format in attacks that typically last only a few hours and can also be launched from zombie machines.
- Email-borne Virus Detection: Like spam and phishing messages, each virus message can be packed differently in terms of its content and the characteristics of the executable files that include the virus. Like spam and phishing, email-borne virus attacks often last for very short durations.

- Inbound vs. outbound detection: The positioning of the detection mechanism also raises additional challenges. Detecting spam as it enters or leaves a network requires different approaches. Inbound traffic is typically accompanied by a high percentage of spam. Outbound traffic typically consists of a majority of legitimate email increasing the chance for false positives (clean email detected as spam). Finding lower volume outbound spam also requires a dedicated approach.

The Cyren approach

The Cyren approach is based on the understanding that all threat outbreaks share some common characteristics, including:

- Most email messages within the outbreak have been altered to make it difficult to set blocking rules based on content analysis.
- Most outbreaks include millions of email messages to maximize the highest possible response rate and the greatest ROI for the attacker. Some attacks though, will be considerably smaller.
- Most outbreaks are released within a short period of time, requiring a real-time solution to detect the outbreak to limit or avoid the damage that can be incurred.
- The originators of the attacks invest heavily in disguising their origin to make it difficult to track the message back to them.

Message patterns

Outbreaks which distribute spam, phishing, and email-borne viruses, consist of messages intentionally composed differently in order to evade commonly-used filters. Nonetheless, all messages within the same outbreak share at least one or more unique, identifiable patterns or values which can be used to distinguish the outbreak. Some examples of these identifiable values:

- In the case of spam the objective is to lead the recipient to the same commercial web sites that can be classified as spam.
- Pseudo-random combinations of the characters from the subject and body of the email will be repeated in an outbreak.
- Different spam attacks are often launched from the same network of zombie machines that can be blacklisted.

- In the case of phishing, the objective is often to lead the victims to the same set of faked URLs.
- Email-borne viruses always contain the same malicious code (otherwise it is a different virus or, more commonly, a variant of the same virus).

All these are recurring values of typical outbreaks. These values are called the 'message patterns' of the outbreak. Any message containing one or more of these unique patterns can be assumed with a great deal of certainty to be part of an outbreak and therefore spam.

Message patterns can be divided into:

- Distribution patterns – the characteristics of the senders (how many, location) and the volume of the emails sent over a period of time.
- Structure patterns – random combinations of text from the header, and body of the message as well as URLs that are found to be repeated in different messages. An example of this approach is shown below. Note that no content analysis is required.

The challenges of message pattern classification include:

- Determining which message patterns identify outbreaks without generating cases of false positives. All outbreaks attempt to disguise messages as legitimate email correspondence pretending to arrive from trusted sources and therefore, solutions that are based on pattern analysis must be able to tell the difference between 'good' and 'bad' patterns and avoid making mistakes.
- Extracting and analyzing these patterns before the outbreak ends. Most outbreaks have a relatively short lifecycle measured in only a few hours. Therefore, any solution that does not detect and classify messages in real-time will only be effective towards the end of the outbreak, when most of the damage has already been done.

The challenges are made more complex by the fact that each new outbreak usually introduces completely new patterns that were not previously analyzed and are therefore unknown to the pattern analyzer. Because spammer tactics are constantly evolving, it is necessary to proactively identify new patterns in real-time in order to determine new outbreaks as they are released.

Recurrent-pattern detection (RPD™) technology

Recurrent Pattern Detection (RPD) technology overcomes the challenges listed above to detect and classify all types of email-borne threat patterns in real-time. RPD is hosted in the Cyren GlobalView™ Cloud, which proactively analyzes billions of Internet transactions daily.

RPD, based on Cyren's U.S. patent #6,330,590, extracts and then analyzes relevant message patterns, which are used to identify email-borne outbreaks. RPD classifies both distribution patterns and structure patterns and the analysis results are stored in a vast database of classifications. In addition to identifying new threat patterns, RPD is also used to modify or enhance classifications of already-identified message patterns. Local instances of RPD are used in the GlobalView™ Cloud to accurately detect low volume or regional Outbound Spam in conjunction with Cyren's Outbound Anti-Spam solution.

Message patterns are extracted from the message envelope, headers, and body with no reference to the message content. RPD therefore has the following additional advantages:

- RPD can be used to identify outbreaks in any language, message format, and encoding type.
- New outbreaks are identified within minutes.
- RPD is designed to distinguish between the patterns of solicited mass emails which represent legitimate business correspondence, such as newsletters, from those of unsolicited spam.
- Cyren uses RPD in a highly scalable environment to deliver extremely high performance rates.
- RPD technology is fully automated and requires no human intervention.
- To ensure maximum privacy and business confidentiality, RPD analyzes hashed values of message patterns and not the 'open' values nor the message content.

RPD identifies nearly 100% of incoming threat messages with almost no cases of false positives. It is language-agnostic and is equally effective for all message formats and encoding types.

SUMMARY

To effectively combat email-borne threats, a successful solution must address a growing number of challenges. Cyren's RPD is a proactive detection technology that continues to outwit those who continue to invent new methods to propagate email-borne threats because it does not rely on the contents of the email and therefore, it is able to detect spam in any language and in every message format including: images, HTML, non-English characters, single and double-byte character sets, etc. RPD technology offers:

- High spam detection rate with almost no cases of false positives
- Early detection of virus threats
- Protection against phishing attempts
- Content-agnostic threat protection
- Multi-language threat detection
- Multi-format threat detection

RPD is an essential technology component among several providing real-time threat data to Cyren's security cloud and OEM services available for fast integration into a wide range of service provider and vendor environments. Because RPD uses future-proof pattern analysis, it also provides the best protection of investment for service providers and vendors of messaging and security applications.